INTRODUCTION TO OPTIMIZATION FOR ECONOMISTS

Introduction to Optimization for Economists

Jacob Engwerda

1st edition
2026

OPEN PRESS TiU

Open Press Tilburg University
Tilburg, The Netherlands

Design and typesetting: LINE UP boek en media bv
Cover design: Bas Ekkers

ISBN: 9789403860336
DOI: https://doi.org/10.56675/9789403860336

Contents

Prologue

This book originates from a crash-course *refresher in mathematics* that I taught for many years at Tilburg University to research master and beginning PhD students in economics, prior to my retirement in 2019. After my retirement, the lecture notes from this course were used as the foundation for writing this book. Additional theory was inorporated, and to make the material accessible at different levels, numerous exercises were added. All exercises are worked out in the accompanying volume *Exercise Solutions for An Introduction to Optimization for Economists*.

A disclaimer is in order. Many of the exercises have not yet been tested by students. Although I have made every effort to work them out carefully, typographical errors or other inaccuracies may still be present. Readers are therefore encouraged to report any errors on the website mentioned below, so that future readers may benefit from these corrections.

My personal hope is that students around the world will enhance their understanding through the mathematical approach to optimization presented in this book. In particular, I believe the book can serve as a useful stepping stone towards the study of more specialized areas in optimization.

I would like to thank Tilburg University, and especially Tilburg University Press, for making it possible to publish this book as open access. I am particularly grateful to Beatriz Lourenço Barrocas Neves Ferreira for her support throughout the publishing process.

Last but not least, I would like to thank my wife, Carine, for giving me the space and freedom to devote so much of my time writing this book and conducting research.

If you have any comments on the book, you are most welcome to post and share them on the DOI-site: https://doi.org/10.56675/9789403860336.

Writing this book has been a rewarding way to spend part of my retirement. If you feel inclined to show your appreciation, you may consider a small donation to the bank account below. I will use this money then to make a hiking tour with Carine. You may use next Wise bank account for that purpose:
Jacob Christiaan Engwerda, IBAN: BE21967500746803, Belgium,
Swift/BIC: TRWIBEB1XXX.

Hoping you may benefit from the book, Jacob Engwerda.

Preface

Economics is the social science that studies the production, distribution, and consumption of goods and services. A central concern of the discipline is how economic agents behave and interact, and how economies function. In line with this focus, a fundamental distinction in economics is made between microeconomics and macroeconomics.

Microeconomics examines the behavior of individual components of the economy, such as households and firms, buyers and sellers, as well as markets and their interactions. Macroeconomics, by contrast, analyzes the economy as a whole and addresses issues such as unemployment, inflation, economic growth, and monetary and fiscal policy.

To support such analyses, economists often employ mathematical tools to model specific situations. By using mathematical abstractions of real-world phenomena, one aims to gain deeper insight into the underlying mechanisms at work. This, in turn, may lead to a better understanding of agent behavior and a more efficient allocation of scarce resources.

In this book, we review a number of standard mathematical tools developed to determine optima of functions. Consequently, the focus lies on the second aspect of the analysis described above: given a specific mathematical model of an economic problem, how can one derive conclusions that - provided the model adequately captures the key features of the problem - contribute to a better allocation of resources?

Starting from the concept of a function, we present conditions under which optima can be derived. Both single-variable and multi-variable functions are considered. Since resources are typically scarce, we then study optimization problems in which the domain of the function is restricted, giving rise to so-called constrained optimization problems. A particularly important subclass of these problems consists of convex optimization problems. Due to their favorable numerical properties - especially in the case of linear programming - they are widely used in applications and are therefore treated in a separate section.

Finally, we consider optimization problems in which outcomes are affected by the actions of more than one agent. This line of research is known as game theory. In the final section of the book, we introduce and study several commonly used concepts from this field.

This textbook is intended for students who wish to acquire mathematical techniques that they can later apply to their own economic problems. The main text is accessible to students of economics as well as those pursuing more technically oriented fields of study.

Mathematical technicalities required for full theoretical rigor are largely omitted from the main text. The objective is to provide students with a clear understanding of the underlying ideas, an intuitive sense of the conditions under which particular mathematical tools can be applied, as well as guidance on how to use them.

Each chapter concludes with a set of exercises, which are divided into three categories. The first category consists of exercises that allow students to test their mastery of the material presented. The second category encourages deeper engagement with mathematical rigor, in particular through proofs of theorems introduced in the main text and related topics. The third category how the presented theory can be applied in economics to gain better insight into real-world phenomena.

All exercises are fully worked out in a separate accompanying volume entitled *Exercise Solutions for An Introduction to Optimization for Economists*.

Finally, the book concludes with a short bibliography, providing students with references to additional literature and related topics.

Notation and Symbols

$f(x) = O(x)$ as $x \to 0$	there exist $c, \ \epsilon$ such that $\|f(x)\| \leq c\|x\|$ when $0 < \|x\| \leq \epsilon$.
$\mathbb{R}$	Real numbers
$\mathbb{R}_+$	strictly positive real numbers
$\mathbb{R}^n$	set of vectors with n real entries
$\mathbb{R}^{n \times m}$	set of $n \times m$ matrices with real entries
$f^{-1}(x)$	inverse function of $f(x)$
$f^{(n)}(x)$	n^{th} order derivative of $f(x)$
x^T	transpose of the vector x
$\|x\|$	length $\sqrt{x^T x}$ of vector x, or, distance vector x to the origin
$\frac{\partial f}{\partial x_i}$, or, $D_i f(x)$, or $f^{'}_{x_i}(x)$	partial derivative of f w.r.t. x_i
$D_u f(x)$	directional derivative $f(x)$ in the direction u
$Df(x) = [D_1 f(x) \cdots D_n(x)]$	row matrix of partial derivatives (or, Jacobian of $f(x)$)
$f^{'}(x)$	the derivative of $f(x)$ (which equals $Df(x)$ if it exists)
C^1	set of differentiable functions which partial derivatives are continuous
C^k	set of functions which all k^{th}-order partial derivatives are continuous
$f^{"}(x)$, or, Hessian $H(x)$	second order derivative of $f(x)$, i.e. $[(D^{'}_1 f(x))^T \cdots (D^{'}_n f(x))^T]^T$
A^T	transpose of matrix A
$A > 0$	positive definite matrix A
$A \geq 0$	positive semi-definite matrix A
$A < 0$	negative definite matrix A
$A \leq 0$	negative semi-definite matrix A
det A	determinant of matrix A
A^{-1}	inverse of matrix A
$I_{n \times n}$	identity matrix in $\mathbb{R}^{n \times n}$
$T_S(a)$	tangent cone to S at a
$TL_S(a)$	set of linearized feasible directions to S at a

$\subset$	subset
$\cup$	union
$\cap$	intersection
$\forall$	for all
$\exists$	there exists
$\mathbb{R}^n/\{0\}$	all $x \in \mathbb{R}^n$ except vector 0
$x \geq y$	componentwise inequality between vectors x and y
$f_D(\lambda)$	function associated with dual problem
$\mathcal{A}$	unit simplex
$S\backslash T$	all elements from S not belonging to T
$\bar{x}_{-i}$	$\bar{x}$, except entry i which is deleted
$\bar{x}_{\backslash i}$	$\bar{x}$, except entry i which is arbitrary
$\mathcal{R}(x_L)$	best response set follower against x_L
Λ	nonnegative orthant

CHAPTER 1

Static Optimization: The Scalar Case

In the first part of this textbook, we study the conditions under which a function of one or more variables attains an optimum. We begin with a detailed analysis of optimization problems involving a single variable. The results are then extended to the multi-variable case in the subsequent chapter.

To identify optima of functions, it is necessary to introduce several preliminary concepts. First, we clarify what is meant by a function; this issue is addressed in the next subsection. In addition, we review a number of elementary functions and some of their key properties that frequently arise in the literature.

A fundamental result in optimization theory is Weierstrass' theorem, which states that a continuous function defined on a compact domain always attains both a global maximum and a global minimum. To gain intuition for this important theorem, we briefly and informally discuss compact sets and continuous functions in two seperate subsections.

1.1 Functions of one variable

We begin by introducing the mathematical concept of a function of a single variable. Let $\mathbb{R}$ denote the set of real numbers.

DEFINITION 1.1 A *function* f from a subset D of $\mathbb{R}$ to $\mathbb{R}$ (write $f : D \to \mathbb{R}$) defines a rule which assigns to each $x \in D$ a *unique* element $y \in \mathbb{R}$. The element y is called the *image* or *value* of the element x and we write $y = f(x)$.
The set D is called the *domain* of f and the set $f(D) := \{f(x) \mid x \in D\}$ the *range* of f.
A point $\bar{x} \in D$ where $f(\bar{x}) = 0$ is called a *zero* of the function f.

Probably the most well-known function is the identity function that returns the value that is used as its argument, i.e., $f(x) = x$. Usually, this function is denoted as $\mathrm{id}(x)$ (so $\mathrm{id}(x) = x$ for all $x \in D$).

EXAMPLE 1.2 Consider $f(x) = x^2 + 1$, for x in the interval $[-1, 1]$. Since to every x in this interval the unique element $x^2 + 1$ is assigned, f defines a function on the interval $[-1, 1]$. The domain of f is the interval $[-1, 1]$ and the range of f is the interval $[1, 2]$.
Note that in this example most of the elements in the range are attained more than once. Furthermore, this function has no zero's, as the equation $x^2 + 1 = 0$ has no solution.

EXERCISE 1.1 Determine for the rules f, defined below, a set D as large as possible such that f defines a function from $D \to \mathbb{R}$. Calculate next the range of these functions.
a. $f(x) = x + 1$. b. $f(x) = \frac{1}{x}$. c. $f(x) = \frac{1}{(x-1)^2}$. d. $f(x) = \{x \mid x^2 = y, \text{ where } y \in [1, 4]\}$.

The *graph* of a function $f(x)$ is a graphic representation of the function in a coordinate system with two axis. The horizontal axis denotes the elements of the domain and the vertical axis the corresponding function value. That is, $\{(x, f(x)) \mid x \in D\}$. Figure 1.1 visualizes the graph of f for Example 1.2.

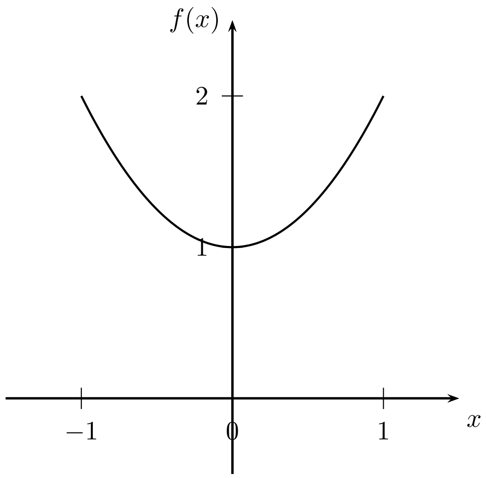


FIGURE 1.1 – *Graph of* $f(x) = x^2 + 1$.

The example illustrates that geometrically we have the following characterization for functions.

PROPOSITION 1.3 *(Vertical line test)*
A rule $f : D \to \mathbb{R}$ defines a function if and only if each vertical line drawn through a point of the domain D meets the graph of f in exactly one point.

In case, geometrically, also each horizontal line drawn through a point y of the range of f meets the graph of f in exactly one point the function is called invertible (or stated in terms of the previous proposition: a horizontal line test applies). The function that expresses the variable x in terms of the variable y is called the inverse function and denoted by $f^{-1}(y)$. More formal,

DEFINITION 1.4 Let $f : D \to \mathbb{R}$ define a function on D with range R. Then, if with each element $f(x) \in R$ there corresponds a unique element $x \in D$ the function f is called *invertible*. Or, stated differently, there exists a function $f^{-1}(y)$ from R to D such that $f(f^{-1}(y)) = y$ for all $y \in R$ and $f^{-1}(f(x)) = x$ for all $x \in D$.
$f^{-1}(y)$ is called then the inverse function of f.

In Exercise 1.30, you are asked to show that if a function is invertible, then it has exactly one inverse function.

EXAMPLE 1.5 Consider the function $f(x) = x^2 + 1$, defined on the interval $[0, 1]$. From Figure 1.1, it is evident that f is invertible on this interval. To determine the inverse function analytically, we solve for x in terms of y from the equation $y = x^2 + 1$. This yields $x^2 = y - 1$. Since x is restricted to be nonnegative on $[0, 1]$, it follows that $x = \sqrt{y-1}$ is the specification we are looking for. Therefore, the inverse function is $f^{-1}(y) = \sqrt{y-1}$. Replacing the symbol y by the symbol x, we obtain $f^{-1}(x) = \sqrt{x-1}$.

EXERCISE 1.2 Determine which of the functions f, defined below, is invertible. For all functions that are invertible, determine the inverse function f^{-1}.
a. $f(x) = x + 1$. b. $f(x) = x^2,\ x \in [-1, 1]$. c. $f(x) = -x^2 + 1,\ x \in [0, 1]$.

Note that geometrically the inverse function is obtained from f by plotting the same graph but now with the y-axis as the horizontal axis and the x-axis as vertical axis. Geometrically this manipulation can be obtained by a reflection of the graph of f through the line $y = x$. Figure 1.2 illustrates this for Example 1.5.

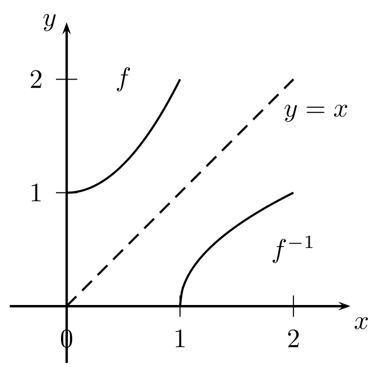


FIGURE 1.2 – *Graph of* $f(x) = x^2 + 1$ *and it's inverse on* $[0, 1]$.

We list some frequently occurring functions below:

A. *Affine functions*: $f(x) = ax + b$, where a, b are scalars.

Geometrically, an affine function represents a straight line, where b is the *intercept* on the y-axis and a is the *slope of the line* (see Figure 1.3).

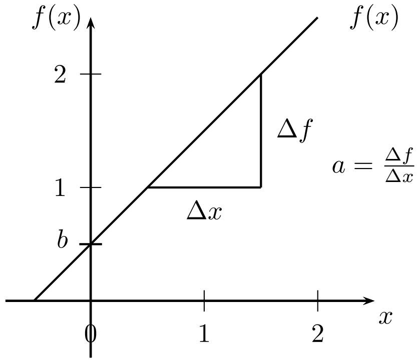


FIGURE 1.3 – *Affine function* $f(x) = ax + b$.

B. *Polynomial functions*: $f(x) = a_0 + a_1 x + \cdots + a_n x^n$, where a_i are scalars.

Here n is called the *degree of the polynomial* (assuming that $a_n \neq 0$). Special cases are $n = 0$, the *constant function*, and $n = 1$, the affine function. The equation $f(x) = 0$ has at most n (real) solutions. Or, stated differently, the function has at most n *zero's*. Figure 1.4 visualizes a typical plot for a third order degree polynomial.

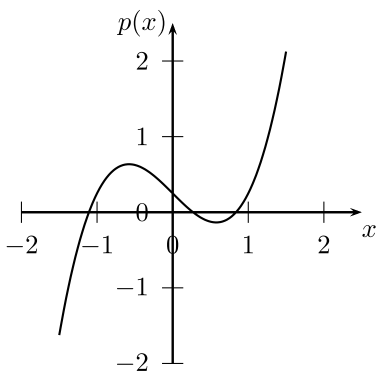


FIGURE 1.4 – *Graph of a third order polynomial* $p(x)$.

C. *Rational functions*: Suppose $p(x)$ and $q(x)$ are polynomial functions. Then $f(x) = \frac{p(x)}{q(x)}$, for all x where $q(x) \neq 0$, is called a rational (ratio of polynomials) function.

Typical for the graph is that at those points where $q(x) = 0$ usually $f(x)$ attains either arbitrarily large values (in which case we say $f(x)$ diverges to ∞ (infinity)) or either arbitrarily small values (in which case we say $f(x)$ diverges to $-\infty$). Figure 1.5 illustrates a typical graph.

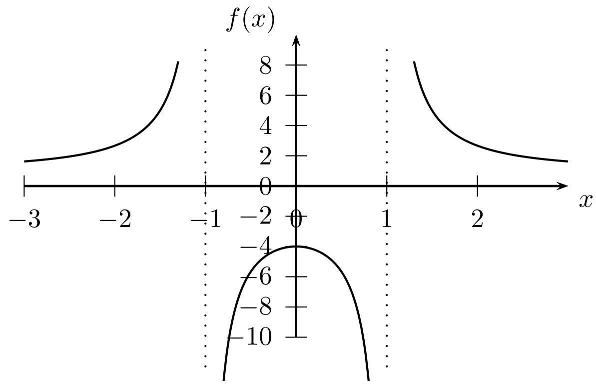


FIGURE 1.5 – *Graph of the rational function* $f(x) = \frac{x^2+4}{x^2-1}$ *for* $x \neq \pm 1$.

D. *Exponential functions*: $f(x) = ma^x$, where a is a positive number and m is some scalar.
The number a is usually referred to as the *base*.
As an example, consider compound interest. Suppose an amount m dollars is invested in a bank that pays an annual compound interest rate of r. The total amount after one year is

$$y(1) = m + mr = m(1 + r).$$

After two years, the accumulated amount is

$$y(2) = m(1 + r) + [m(1 + r)]r = m(1 + r)^2.$$

In general, after x years the accumulated amount is

$$y(x) = m(1 + r)^x.$$

Thus, $y(x)$ is an exponential function with base $(1 + r)$. Figure 1.6 illustrates the graph of an exponential function with base 2.
A particularly important base in the study of exponential functions is Euler's number

$$e = 2.718....$$

This constant arises, for instance, as the limit obtained when a bank pays an interest $r = 1$, and compounds the interest n times per year, with each payment equal to a fraction $\frac{1}{n}$, as n tends to infinity. Formally,

$$e = (1 + \frac{1}{n})^n, \text{ if } n \text{ diverges to infinity.}$$

In particular, this implies that if a bank pays interest at rate r according to this compounding scheme, the accumulated amount after one year is e^r. So, when an exponential function of the form $f(x) = e^x$ appears in the literature, the symbol e refers to this constant.

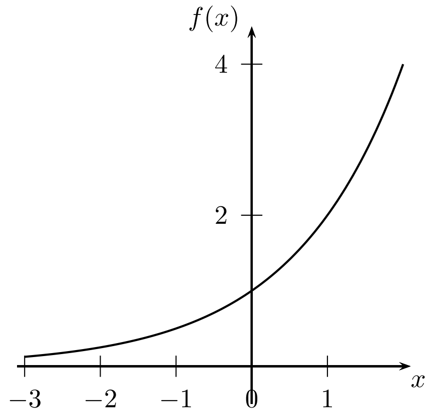


FIGURE 1.6 – *Graph of the exponential function* $f(x) = 2^x$.

E. *Logarithmic functions*: $f(x) = {}^a\log(x)$, where a is a positive number (the *base* of the logarithm)[1]. Using the horizontal line test for the exponential function $y = a^x$, it is clear that this function is invertible. The inverse of this exponential function is called the logarithmic function and denoted by ${}^a\log(y)$. So, by definition

$$x = {}^a\log(y) \text{ if and only if } y = a^x.$$

Since the range of $f(x) = a^x$ is the set $\{y \mid y > 0\}$, this relationship implies that the logarithmic function is defined only for all strict positive scalars.
Special notations are attached to the inverse functions corresponding with the basis 10, Euler's number e, and 2. The logarithm to base $a = 10$ is called the *common logarithm* ($\log(x)$) and has many applications in science and engineering. The *natural logarithm* ($\ln(x)$) has the constant e as its base; its use is widespread in pure mathematics, especially calculus, and economics. The *binary logarithm* uses base $a = 2$ and is prominent in computer science.
Since the graph of the inverse function is obtained by reflecting the graph of the original across the line $y = x$, a typical graph of a logarithmic function appears as shown in Figure 1.7.

A number of elementary properties which immediately follow from the definition of the logarithmic function $f(x) = {}^a\log(x)$ are:
1. $f(0) = 1$.
2. The function is an increasing function (i.e., it rises to the right) for all $a > 1$.
3. The function is a decreasing function (i.e., it falls to the right) for all $0 < a < 1$.

1 Another notation frequently used in literature is $f(x) = \log_a(x)$.

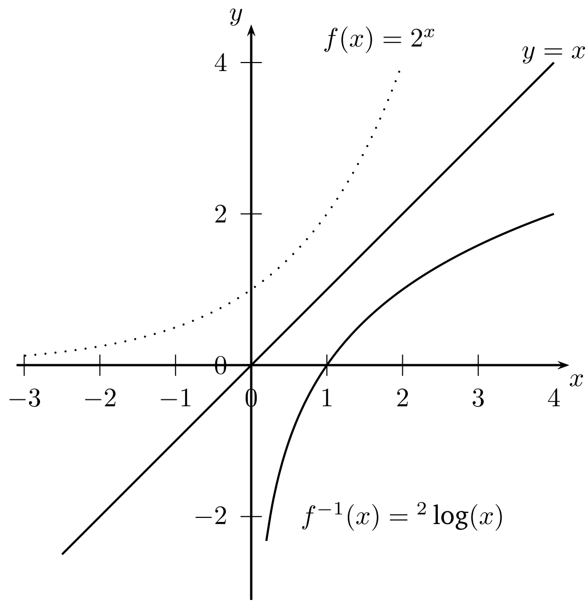


FIGURE 1.7 – *Graph of the logarithmic function* ${}^2\log(x)$.

4. The function is negative when $0 < x < 1$ and $a > 1$.
5. The function is positive when $x > 1$.
6. The function is not defined when x is negative.

Similarly, the following elementary rules for calculation with logarithms result immediately from the rules for calculation with exponents:

Calculation rules for Exponents	Calculation rules for Logarithms
$a^x a^z = a^{x+z}$	${}^a\log(uv) = {}^a\log(u) + {}^a\log(v)$
$\frac{a^x}{a^z} = a^{x-z}$	${}^a\log(\frac{u}{v}) = {}^a\log(u) - {}^a\log(v)$
$(a^x)^z = a^{xz}$	${}^a\log(u^n) = n * {}^a\log(u)$
$a^x = a^y$ implies $x = y$	${}^a\log(x) = {}^a\log(y)$ implies $x = y$

F. *Combining functions*:
Let f and g be two functions from a set D to $\mathbb{R}$. The *sum* of f and g, denoted by $f + g$ is the function from D to $\mathbb{R}$ defined by

$$(f + g)(x) := f(x) + g(x) \text{ for all } x \in D.$$

Similarly, for any real number λ, the function λf is defined by

$$(\lambda f)(x) := \lambda f(x) \text{ for all } x \in D.$$

More generally, the *product* and *quotient* of functions are defined by

$$fg(x) := f(x)g(x) \text{ and } \left(\frac{f}{g}\right)(x) := \frac{f(x)}{g(x)}, \text{ whenever } g(x) \neq 0.$$

Finally, the *composition* of functions is defined as follows. Let: $g : S \to T$ and $f : T \to \mathbb{R}$. Then the composite function $f \circ g$ is given by

$$(f \circ g)(x) := f(g(x)) \text{ for all } x \in S.$$

EXAMPLE 1.6 Let $f(x) = x + 2$ and $g(x) = x^2 + 1$. Then:
$(f + g)(x) = x^2 + x + 3$; $(fg)(x) = (x + 2)(x^2 + 1)$; $\left(\frac{f}{g}\right)(x) = \frac{x+2}{x^2+1}$; $(f \circ g)(x) = x^2 + 3$ and $(g \circ f)(x) = (x + 2)^2 + 1$.

EXERCISE 1.3

a. Simplify $\frac{{}^a\log(x)}{{}^a\log(y)+{}^a\log(z)}$.

b. Simplify $e^{2\ln(x)} - \ln(e^{2x})$.

c. Solve $2^{x^2} = 4^{-3x-4}$.

d. Solve $\left(\frac{1}{4}\right)^{x^2} = 64\left(2^{-2x}\right)^4$.

e. Solve $2\ln(x - \frac{1}{2}e) - \ln(x - \frac{3}{4}e) = 1$.

EXERCISE 1.4 Let $f(x) = x^2 + 1$ and $g(x) = \log(x^2 + 1)$. Determine

a. $(f + g)(x)$. b. $(fg)(x)$. c. $\left(\frac{f}{g}\right)(x)$. d. $(f \circ g)(x)$. e. $(g \circ f)(x)$.

1.2 Compact Sets

To introduce the notion of compactness of a set $S \subset \mathbb{R}$ we first introduce the notions of interval and bounded sets.

DEFINITION 1.7 An *interval* I is a set of real numbers with the property that with every two numbers $x, y \in I$ also every number between them belongs to I.

DEFINITION 1.8 A set $S \subset \mathbb{R}$ is called *bounded* if there exists a number m such that for all $x \in S$ it holds: $|x| < m$.

In describing intervals we use the following notation:

$$\begin{aligned} (a,b) &:= \{x \in \mathbb{R} \mid a < x < b\}; \; [a,b] := \{x \in \mathbb{R} \mid a \leq x \leq b\}; \; [a,b) := \{x \in \mathbb{R} \mid a \leq x < b\}; \\ (a,b] &:= \{x \in \mathbb{R} \mid a < x \leq b\}. \end{aligned}$$

These are all bounded intervals. Figure 1.8, panel a, illustrates the bounded interval $[1,2]$. We will also consider unbounded intervals. For these we use the notation:

$$\begin{aligned}(a,\infty) &:= \{x \in \mathbb{R} \mid x > a\}; \;\; [a,\infty) := \{x \in \mathbb{R} \mid x \geq a\}; \;\; (-\infty,b) := \{x \in \mathbb{R} \mid x < b\}; \\ (-\infty,b] &:= \{x \in \mathbb{R} \mid x \leq b\}.\end{aligned}$$

Next we introduce the notion of openness and closedness of a set S. To that end first recall that the *complement set* of a set S are all numbers that don't belong to S, i.e. $\{x \in \mathbb{R} \mid x \notin S\}$. So, for example, if $S = \{1,2\}$ then the complement set of S are all numbers in $\mathbb{R}$ different from 1 and 2.

DEFINITION 1.9 A set S of real numbers is called *open* if for every $x \in S$ there exists an interval (a_x, b_x) containing x, such that all elements of this interval belong to S.
The set S is called *closed* if the complement set of S is open. Or, equivalently[2], if one considers an arbitrary sequence of numbers $\{s_n\}_{n=1}^{\infty}$, all belonging to S, that converges to some limit, s, then this limit s also belongs to S.

(a) Compact interval $\{x \mid 1 \leq x \leq 2\}$.

(b) Bounded set S: $\{x \mid -1 < x \leq 0 \, \cup \, 1 \leq x \leq 2\}$, neither open or closed. Accumulation points S: $\{-1,0,1,2\}$. Boundary points S: $\{0,1,2\}$.

FIGURE 1.8 – *Sets and properties.*

With this definition the intervals (a,b), (a,∞) and $(-\infty,b)$ are open, the intervals $[a,b]$, $[a,\infty)$ and $(-\infty,b]$ are closed, the intervals $[a,b)$ and $(a,b]$ are neither open or closed and the set of all real numbers, $\mathbb{R}$, is as well open as closed (see Exercise 1.28).
A simple yet often useful property of a finite collection of open (or closed) sets is that both their union and intersection share the same property. We formalize this property below.

PROPOSITION 1.10 *If S and T are two open sets, then both $S \cap T$ and $S \cup T$ are open as well. Similarly, if S and T are closed sets, then $S \cap T$ and $S \cup T$ are also closed .*
More generally, the intersection of an arbitrary (possibly infinite) collection of closed sets is closed, and the union of an arbitrary (possibly infinite) collection of open sets is open.

Using the above definitions we can define finally the notion of a compact set.

2 This can be proved formally, but is not trivial to show.

DEFINITION 1.11 A set S of real numbers is called *compact* if it is both bounded and closed.

It is easy to verify that in Figure 1.8, panel (a), the interval $[1,2]$ is compact.
The set S in panel (b), which is the union of the intervals $(-1,0]$ and $[1,2]$, is bounded but not compact, since S is not closed.
To see this formally, consider the sequence $\{s_n\}_{n=1}^{\infty}$ defined by $s_n = -1 + \frac{1}{n}$. This sequence converges to -1. Notice that $s_n \in S$ for all $n = 1, 2, ...$ Therefore, -1 is an accumulation point of S, yet $-1 \notin S$.

EXAMPLE 1.12 Consider $S = (1,2] \cup [3,4]$. Then S is bounded, not open, not closed and not compact.
Consider $S = [1,2] \cup [3,4]$. Then S is bounded, not open, closed and compact.
Consider $S = [1,2] \cup [3,\infty)$. Then S is not bounded, not open, closed and not compact.
Consider $S = \{1\}$. Then S is bounded, not open, closed and compact.

Another intuitive way to formalize the concept of a set being closed is by examining its boundary and its accumulation points.

DEFINITION 1.13 A point $b \in S$ is called a *boundary point* of the set S if every interval (b_l, b_r), containing b, contains at least one point not belonging to S. Or, equivalently, if there exists a sequence of numbers, all not belonging to S, that converges to b.
A point a is called an *accumulation point* of the set S if every interval (a_l, a_r) containing a, contains at least one point in S different from a and at least one point not in S.
Equivalently, a is an accumulation of S if there exists a sequence of points not in S that converges to a, and another sequence of points in S (with all terms distinct from a beyond a certain index) that also converges to a.

Notice the subtle difference between an accumulation point and boundary point of a set S, that the first-mentioned one does not necessarily have to belong to the set, whereas a boundary point by definition always belongs to the set S.

EXAMPLE 1.14 Consider the interval $I = [0,1)$. I has one boundary point: 0. The accumulation points of I are 0 and 1.

By definition 1.9, we conclude that a set S is closed if and only if it contains all of its accumulation points. Or, stated differently, all its accumulation points are boundary points.

EXERCISE 1.5 Determine the boundary and accumulation points of next sets. Motivate which of these sets is bounded, open, closed or compact, respectively.
a. $[1,2] \cup [3,\infty)$. b. $\{1,2\}$. c. $[1,2) \cup (2,3]$. d. $[1,3) \cap (0,2]$.

1.3 Continuous Functions

In this section we introduce in an informal way the notion of continuity of a function. A formal set-up is skipped. Readers interested in a more formal approach of this notion can find some more background on this in the exercises (see, e.g., Exercise 1.32). Otherwise, for a more extensive treatment, they should consult the literature (see, e.g., [6][Section 6.4]). This section also provides some important properties of continuous functions and we argue that most functions encountered in literature are continuous.

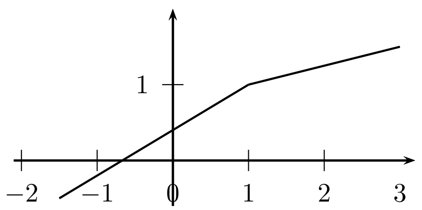
Intuitively, a function is called continuous at a point x_0 if its graph has no "jump" or "break" above x_0. In other words, the graph is connected within an open interval containing x_0, so that it can be drawn without lifting the pencil from the paper.

Equivalently, a function f is continuous at x_0 if, for any sequence $\{s_n\}_{n=1}^{\infty}$ in the domain that converges to x_0, the corresponding sequence of function values $\{f(s_n)\}_{n=1}^{\infty}$ converges to $f(x_0)$.

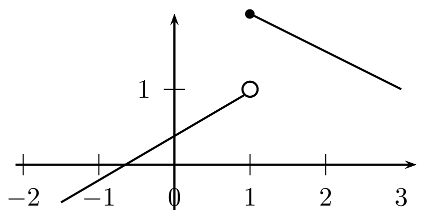
DEFINITION 1.15 Let the function f be defined on an open interval I and $x_0 \in I$. Then f is called *continuous at* $x = x_0$, in case for every sequence $\{s_n\}_{n=1}^{\infty}$ that approaches x_0 it holds that $\{f(s_n)\}_{n=1}^{\infty}$ approaches $f(x_0)$. If f is continuous at every $x_0 \in I$, f is called *continuous* on I. If f is not continuous it is called *discontinuous*.

Clearly this definition can be straightforwardly generalized to define continuity of a function that is defined on a set S that is the union of open intervals that have an empty intersection. Figure 1.9 illustrates the basic idea of continuity. Moreover, Figure 1.5 illustrates a continuous function which is defined on a set which is the union of disjunct intervals.

In the sequel we will make the convention that if we talk about continuity of a function on a set S, S is the union of open intervals that have an empty intersection.



(a) A continuous function.



(b) A discontinuous function.

FIGURE 1.9 – *Continuity of functions.*

Using this formalization it can be easily shown that next theorem applies.

THEOREM 1.16 *Let f and g be continuous functions on a set S, and λ, μ be real numbers. Then the following functions are continuous on S too:*

$$i)\ \lambda f + \mu g;\ ii)\ fg;\ iii)\ \frac{f}{g},\ \text{provided that } g(x) \neq 0 \text{ for all } x \in S.$$

Furthermore, if f is continuous on the range of $g(S)$ and g is continuous on S, also the composite function $f \circ g$ is continuous on S.

Clearly the constant functions and the linear function $f(x) = x$ are continuous on $\mathbb{R}$ (see also Exercise 1.32). Using repeatedly Theorem 1.16, items i) and ii), shows then that every polynomial function is continuous too. Consequently, using item iii), it follows that also all rational functions are continuous on all intervals where they are defined. It can also be shown that the exponential functions are continuous on $\mathbb{R}$ and the logarithmic functions, for $x > 0$, too. So, by combining all these functions using addition, multiplication, division (except by zero), and composition we conclude from Theorem 1.16 that such functions are continuous at all points where they are defined.
We now come to a result which may seem so obvious as to be hardly worth mentioning. It is, however, of fundamental importance and serves as a foundation stone in the development of calculus.

PROPOSITION 1.17 *(Continuity property)*
Let the continuous function f be defined on an interval I. Then,
1. *the set $f(I)$*[3] *is also an interval;*
2. *if I is a compact interval $f(I)$ is also a compact interval.*

Figure 1.10 visualizes Proposition 1.17. One important corollary of this result is the next so-called Intermediate Value Theorem.

THEOREM 1.18 *(Intermediate Value Theorem)*
Let the continuous function f be defined on an interval I. Assume $x_i \in I,\ i = 1, 2$, with $x_1 < x_2$. Then, if y lies between $f(x_1)$ and $f(x_2)$ there exists an x_3 between x_1 and x_2 such that $f(x_3) = y$.

From Proposition 1.17 one can also derive immediately next important theorem in optimization.

THEOREM 1.19 *(Weierstrass Theorem)*
Let f be continuous on a compact set S. Then, f attains both a maximum value and a minimum value on S.

3 formal this set is defined as $\{f(x) \mid x \in I\}$.

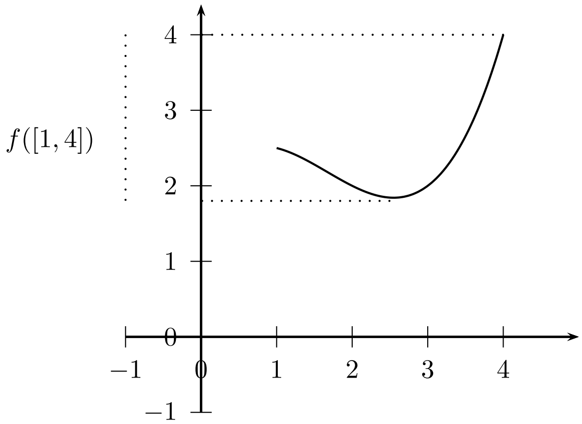


FIGURE 1.10 – *Range of continuous function of compact set.*

EXAMPLE 1.20 Consider the function $f(x) = x^2 - xe^{x^2-4x+1} + \ln(x^2+1)$ on $[0,1]$. From Theorem 1.16 it follows straightforwardly that f is continuous on $[0,1]$. Furthermore, $[0,1]$ is a compact set. So by Weierstrass' theorem, f attains both a maximum and minimum on $[0,1]$.
Notice (as we will see later on) that it is analytically not possible to calculate the precise locations where this maximum and minimum are attained.

EXERCISE 1.6 Motivate which of the next functions is continuous.

a. $f(x) = \begin{cases} \log(x^2+1), & x \le 0 \\ x, & x > 0 \end{cases}$.

b. $f(x) = xe^{x^2+\ln(x^2+1)}$.

c. $f(x) = \begin{cases} \frac{1}{x}, & x \neq 0 \\ 0, & x = 0 \end{cases}$.

EXERCISE 1.7 Motivate whether the next functions have an optimum.

a. $f(x) = \frac{1}{x}$, $x > 0$.

b. $f(x) = \frac{x^2-e^x}{\ln(x^2+2)}$, $-10 \le x \le 12$.

1.4 Differentiable Functions

The rate at which a function changes - i.e., the change in $f(x)$ per unit change in x - is a fundamental concept. It allows us to approximate how much the value of a function will increase or decrease when x changes by one unit. In economics, this rate of change is referred to as the *marginal value*. For instance, the marginal production cost measures how total production costs change when an input factor is increased by one unit.

At a given point x_0, this rate of change can be approximated by the so-called derivative of the function at x_0, provided the function behaves smoothly. At points where a function reaches a maximum or a minimum, this rate of change equals zero. Consequently, derivatives play a central role in solving optimization problems. By determining all points at which the derivative equals zero, one obtains an initial set of candidate points where an optimum may occur.

For the elementary functions introduced earlier, in fact, derivatives can be computed analytically. By assigning to every point in the domain of the function its corresponding derivative value, we obtain a function again, the so-called derivative function. So, whenever a function behaves smoothly, solving an optimization problem essentially reduces to finding the zeros of the derivative function.

This leads to the question of what it means for a function to be *smooth*. From a geometric perspective, a function f is said to be smooth at a point x_0 if there exists a unique line that just "touches" the graph of f at that point. When this is the case, function values near x_0 can be accurately approximated using this line, allowing us to reliably estimate the effect of small changes in x on the function value. This idea is illustrated in Figure 1.11. Informal, a line that just "touches" the graph of f at the point x_0 is called *a tangent line* to the graph of f at x_0.

So, the question is therefore under which conditions there will be a unique tangent line to the graph of f at x_0. To answer that question we consider the average change in the function value if the argument of the function changes from x_0 to $x_0 + \Delta x$. This change is represented by the so-called *difference quotient*:

$$\frac{\Delta f(x_0)}{\Delta x} := \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}.$$

For example consider $f(x) = 3x^2 - 4$. Then the average change in f that occurs at x_0 if x changes into $x_0 + \Delta x$ is given by evaluating f at x_0 and at $x_0 + \Delta x$, respectively. Notice that $f(x_0) = 3x_0^2 - 4$ and $f(x_0 + \Delta x) = 3(x_0 + \Delta x)^2 - 4$. Therefore, the average change is

$$\frac{\Delta f(x_0)}{\Delta x} = \frac{3(x_0 + \Delta x)^2 - 4 - [x_0^2 - 4]}{\Delta x} = \frac{6x_0\Delta x + 3(\Delta x)^2}{\Delta x} = 6x_0 + 3\Delta x.$$

So, if $x_0 = 3$ and $\Delta x = 4$, the average change of f is $6 * 3 + 3 * 4 = 30$. This means that, on the average, as x changes from 3 to 7, the change in f is 30 units per unit change in x.

Figure 1.11 visualizes this difference quotient. From the graph we infer that geometrically the difference quotient measures the slope of the chord PQ, i.e. $\frac{\Delta f(x_0)}{\Delta x} = \frac{QR}{PR}$. Furthermore we observe that if Δx goes to zero, the slope of the chord PQ approaches the slope of the line m, which is tangent to $f(x)$ in the point x_0. Note that in this graph we took Δx positive, but that this property also holds if we would have chosen Δx negative. That is, if we would have approached $f(x_0)$ from the left.

By letting Δx going to zero in the difference quotient $\frac{\Delta f(x_0)}{\Delta x}$ we measure in fact the instantaneous effect a change in x_0 will have on $f(x)$. In the sequel we will call this limit value (if it exists) the *derivative* of f at the point x_0 and denote it by $f^{'}(x_0)$. In case this derivative exists, we will call the line through $f(x_0)$ which slope equals this derivative *the tangent line* of f at x_0. The equation of this line is: $m(x) = f(x_0) + f^{'}(x_0)(x - x_0)$.

In fact, using this equation for the tangent line, the statement that the difference quotient $\frac{\Delta f(x_0)}{\Delta x}$ approaches zero as Δx approaches zero can also be reformulated in terms of our original problem

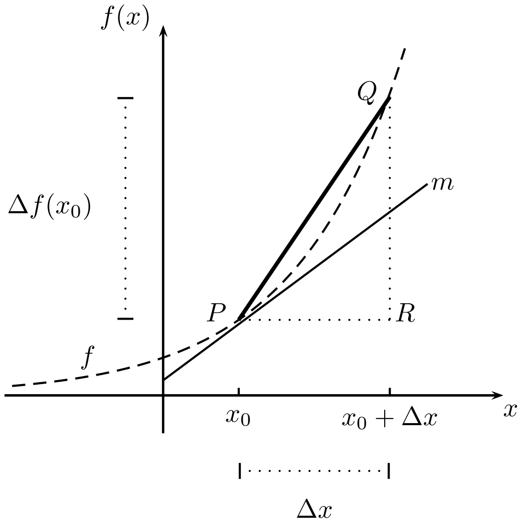


FIGURE 1.11 – *Difference quotient and differentiability.*

formulation. That is, notice that:

$$\frac{\Delta f(x_0)}{\Delta x} = f^{'}(x_0) \text{ if } \Delta x \text{ approaches zero} \Leftrightarrow$$
$$\frac{f(x_0 + \Delta x) - f(x_0) - f^{'}(x_0)\Delta x}{\Delta x} = 0 \text{ if } \Delta x \text{ approaches zero} \Leftrightarrow \text{(Replace } \Delta x = x - x_0)$$
$$\frac{f(x) - f(x_0) - f^{'}(x_0)(x - x_0)}{x - x_0} = 0 \text{ if } x \text{ approaches } x_0 \Leftrightarrow$$
$$\frac{f(x) - m(x)}{x - x_0} = 0 \text{ if } x \text{ approaches } x_0.$$

From the last equation, we infer that existence of the derivative $f^{'}(x_0)$ is equivalent with next statement. The distance between the function value $f(x, y)$ and the corresponding value on the tangent line, $m(x) = f(x_0) + f^{'}(x_0)(x - x_0)$, divided by the distance of x to x_0, goes to zero if x goes to x_0. This formulation will be used in the next chapter to generalize the notion of ”smoothness” to functions of more than one variable.
Next definition summarizes in a formal way the above discussion.

DEFINITION 1.21 Suppose f is defined on an open interval I containing the point x_0. The function f is said to be *differentiable* at x_0 if the difference quotient

$$\frac{\Delta f(x_0)}{\Delta x}$$

has a unique limit as Δx approaches zero. This limit is called the *derivative* of f at x_0 and is denoted by $f^{'}(x_0)$ or $\frac{df}{dx}(x_0)$.
The *tangent line* to the graph of f at x_0 is defined as the line passing through the point $(x_0, f(x_0))$

with slope $f^{'}(x_0)$.
If f is differentiable at every point of I, then f is said to be differentiable on I.

REMARK 1.22 As already notice above, the tangent line to the graph of f is given by

$$m(x) = f(x_0) + f^{'}(x_0)(x - x_0),$$

if f is differentiable at x_0.
Usually, the tangent line at x_0 has the intuitive property of ”touching” the graph of f at that point: that is, there exists an open interval I containing x_0 such that the tangent line does not intersect the graph of f anywhere else within I. However, this is not always the case.
This phenomenon is illustrated in Section 1.8. In Example 1.51, we consider a function f that is differentiable on $\mathbb{R}$, with $f^{'}(0) = 0$ and $f(0) = 0$, yet has the property that, in every interval $(x_l, 0)$ with $x_l < 0$, f attains both positive and negative values. Consequently, the tangent line at $x_0 = 0$ given by $m(x) = 0$, intersects the graph of f on any open interval to the left of $x = 0$.

In the example considered above, the difference quotient for the function $f(x) = 3x^2 - 4$ at x_0 is

$$\frac{\Delta f(x_0)}{\Delta x} = 6x_0 + 3\Delta x.$$

As Δx goes to zero, this difference quotient approaches the value $6x_0$. So, the derivative of $f(x)$ at x_0 is

$$f^{'}(x_0) = 6x_0.$$

The tangent line to the graph of f at, for instance, $x_0 = 1$ is: $y(x) = -1 + 6(x - 1)$.

From the definition it is clear that not every function is differentiable. A well-known example to illustrate this is the function $f(x) = |x|$. We plotted the graph of this function in Figure 1.12. Obviously this graph has not a unique tangent line at $x_0 = 0$. So this function is not differentiable at $x_0 = 0$. This intuition is confirmed by considering the difference quotient. $\frac{\Delta f(x_0)}{\Delta x} = -1$ if $\Delta x < 0$ and $\frac{\Delta f(x_0)}{\Delta x} = 1$ if $\Delta x > 0$. Therefore, the difference quotient does not approach a unique value if Δx approaches zero.
It can be shown from the definition 1.21 (see Exercise 1.33) that if a function is differentiable at a point x_0 the function must be continuous at this point too. Or, stated differently, a prerequisite for a function to be differentiable at a point x_0 is that it is continuous at that point. We stress this point by formulating next theorem.

THEOREM 1.23 *Let f be differentiable on an open interval containing x_0. Then f is continuous at x_0 too.*

As with continuity, the following property can be established for differentiable functions.

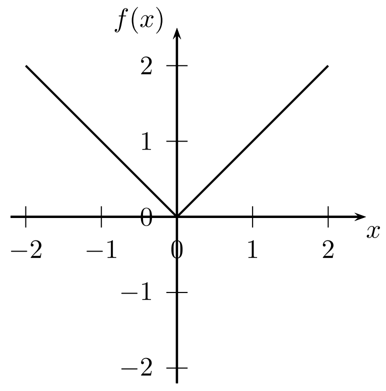


FIGURE 1.12 – $f(x) = |x|$.

THEOREM 1.24 *Let f and g be differentiable on an open interval containing x_0, and λ, μ be real numbers. Then*
1. $\lambda f + \mu g$ *is differentiable at* x_0, *and* $(\lambda f + \mu g)^{'}(x_0) = \lambda f^{'}(x_0) + \mu g^{'}(x_0)$
2. fg *is differentiable at* x_0, *and* $(fg)^{'}(x_0) = f(x_0)g^{'}(x_0) + f^{'}(x_0)g(x_0)$
3. $\frac{f}{g}$ *is differentiable at* x_0 *(provided* $g(x_0) \neq 0$*), and* $\left(\frac{f}{g}\right)^{'}(x_0) = \frac{f^{'}(x_0)g(x_0) - g^{'}(x_0)f(x_0)}{g^2(x_0)}$
Moreover, if f is differentiable at $g(x_0)$ and g is differentiable at x_0, also
4. $f \circ g$ *is differentiable at* x_0, *and* $(f \circ g)^{'}(x_0) = f^{'}(g(x_0))g^{'}(x_0)$.

It is easily verified from Definition 1.21 that the constant function is differentiable and its derivative is zero at any point. It is also clear that the linear function $f(x) = x$ is differentiable everywhere and its derivative is 1 at any point. By a repeated use of Theorem 1.24 it follows then that every polynomial function is differentiable too and, consequently, also every rational function (at points where the denominator differs from zero). In fact one can conclude relatively easily using the above arguments that the derivative of $f(x) = x^n$ is $f^{'}(x) = nx^{n-1}$. Geometrically it is also clear that any exponential function and logarithmic function are differentiable. This implies that most of the functions that occur in literature are differentiable.
In Table 1, below, we list the derivatives of some elementary functions.
Notice from this table that the derivative of an elementary function is an elementary function again. This implies that most of the derivatives of the functions that occur in literature are differentiable again, etc. Or, stated differently, most of the functions can be differentiated as often as we want.

EXAMPLE 1.25 From Theorem 1.24 and Table 1, it is immediately clear that the following two functions are differentiable. Moreover, their corresponding derivatives follow directly.
1. If $f(x) = \frac{1}{x} + x^2 \ln(x)$, $x > 0$, $f^{'}(x) = -\frac{1}{x^2} + 2x\ln(x) + x^2\frac{1}{x} = -\frac{1}{x^2} + 2x\ln(x) + x$.
2. If $f(x) = xe^{x^2+1}$, $f^{'}(x) = e^{x^2+1} + xe^{x^2+1}2x = (1 + 2x^2)e^{x^2+1}$.